

# Uniform Approximation and Bracketing Properties of VC classes

Terrence M. Adams <sup>\*</sup> and Andrew B. Nobel <sup>†</sup>

July 2010

## Abstract

We show that the sets in a family with finite VC dimension can be uniformly approximated within a given error by a finite partition. Immediate corollaries include the fact that VC classes have finite bracketing numbers, satisfy uniform laws of averages under strong dependence, and exhibit uniform mixing. Our results are based on recent work concerning uniform laws of averages for VC classes under ergodic sampling.

---

<sup>\*</sup>Terrence Adams is with the Department of Defense, 9800 Savage Rd. Suite 6513, Ft. Meade, MD 20755

<sup>†</sup>Andrew Nobel is with the Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599-3260. Email: nobel@email.unc.edu

# 1 Introduction

Let  $\mathcal{X}$  be a complete separable metric space with Borel sigma field  $\mathcal{S}$ , and let  $\mathcal{C} \subseteq \mathcal{S}$  be a family of measurable sets. For each finite set  $D \subseteq \mathcal{X}$ , let  $\{C \cap D : C \in \mathcal{C}\}$  be the collection of subsets of  $D$  induced by the members of  $\mathcal{C}$ . The family  $\mathcal{C}$  is said to be a Vapnik-Chervonenkis (VC) class if there is a finite integer  $k$  such that

$$|\{C \cap D : C \in \mathcal{C}\}| < 2^k \text{ for every } D \subseteq \mathcal{X} \text{ with } |D| = k. \quad (1)$$

Here and in what follows  $|\cdot|$  denotes cardinality. The smallest  $k$  for which (1) holds is known as the VC-dimension of  $\mathcal{C}$ . Classes of sets having finite VC-dimension play a central role in the theory of machine learning and empirical processes (*c.f.* [7, 9, 4, 5]).

## 1.1 Principal Result

Let  $\mu$  be a probability measure on  $(\mathcal{X}, \mathcal{S})$ , and let  $\pi$  be a finite, measurable partition of  $\mathcal{X}$ . For every set  $C \in \mathcal{C}$ , the  $\pi$ -boundary of  $C$ , denoted  $\partial(C : \pi)$ , is the union of all the cells in  $\pi$  that intersect both  $C$  and its complement with positive probability. Formally,

$$\partial(C : \pi) = \cup \{A \in \pi : \mu(A \cap C) > 0 \text{ and } \mu(A \cap C^c) > 0\}.$$

Note that  $\partial(C : \pi)$  depends on  $\mu$ ; this dependence is suppressed in our notation. Of interest here is the existence of a fixed finite partition  $\pi$  such that the measure of the boundary  $\partial(C : \pi)$  is small for every set  $C$  in  $\mathcal{C}$ . In general, the existence of a uniformly approximating partition depends on the family  $\mathcal{C}$  and the measure  $\mu$ . Our main result shows that VC classes possess this uniform approximation property, regardless of the measure  $\mu$ .

**Theorem 1.** *Let  $\mu$  be a probability measure on  $(\mathcal{X}, \mathcal{S})$ . If  $\mathcal{C}$  is a VC-class, then for every  $\epsilon > 0$  there exists a finite measurable partition  $\pi$  of  $\mathcal{X}$  such that*

$$\sup_{C \in \mathcal{C}} \mu(\partial(C : \pi)) < \epsilon. \quad (2)$$

Several corollaries of Theorem 1 are discussed in the next section. The proof of Theorem 1 is presented in Section 3.

## 2 Corollaries of Theorem 1

Here we present several immediate corollaries of Theorem 1 that may be of independent interest.

## 2.1 Bracketing of VC Classes

Let  $\mu$  be a probability measure on  $(\mathcal{X}, \mathcal{S})$ . For each pair of sets  $A, B \in \mathcal{S}$ , the bracket  $[A, B]$  consists of all those sets  $C \subseteq \mathcal{X}$  such that  $A \subseteq C \subseteq B$ . If  $A$  is not a subset of  $B$ , then  $[A, B]$  is empty. The bracket  $[A, B]$  is said to be an  $\epsilon$ -bracket if  $\mu(B \setminus A) \leq \epsilon$ . The bracketing number  $N_{[\cdot]}(\epsilon, \mathcal{C}, \mu)$  of a family  $\mathcal{C} \subseteq \mathcal{S}$  is the least number of  $\epsilon$ -brackets needed to cover  $\mathcal{C}$ . Note that the sets defining the minimal brackets need not be elements of  $\mathcal{C}$ .

**Corollary 1.** *Let  $\mu$  be any probability measure on  $(\mathcal{X}, \mathcal{S})$ . If  $\mathcal{C}$  is a countable VC-class, then  $N_{[\cdot]}(\epsilon, \mathcal{C}, \mu)$  is finite for every  $\epsilon > 0$ .*

**Remark:** Using routine arguments, the assumption that  $\mathcal{C}$  is countable can be replaced by the weaker assumption that there exists a countable sub-family  $\mathcal{C}_0 \subseteq \mathcal{C}$  such that the indicator function of every set in  $\mathcal{C}$  is the pointwise limit of the indicator functions of sets in  $\mathcal{C}_0$ .

**Proof:** Fix a probability measure  $\mu$  and  $\epsilon > 0$ . Let  $\pi = \{A_1, \dots, A_m\}$  be a finite measurable partition of  $\mathcal{X}$  such that (2) holds, and assume without loss of generality that each set  $A_j$  has positive  $\mu$ -measure. Let  $A_j$  be an element of  $\pi$ . For each  $C \in \mathcal{C}$ , remove points in  $C$  from  $A_j$  if  $\mu(A_j \cap C) = 0$ , and remove points in  $C^c$  from  $A_j$  if  $\mu(A_j \cap C^c) = 0$ . Denote the resulting set by  $B_j$ . Clearly  $B_j \subseteq A_j$  and, as  $\mathcal{C}$  is countable,  $\mu(A_j \setminus B_j) = 0$ . The definition of  $B_j$  ensures that for each  $C \in \mathcal{C}$  exactly one of the following relations holds:  $B_j \subseteq C$ ,  $B_j \subseteq C^c$ , or  $\mu(B_j \cap C) \cdot \mu(B_j \cap C^c) > 0$ . Let  $B_0 = \mathcal{X} \setminus \cup_{j=1}^m B_j$ , and define the partition  $\pi' = \{B_0, B_1, \dots, B_m\}$ . Given  $C \in \mathcal{C}$  let  $C_l = \cup\{B \in \pi' : B \subseteq C\}$  and  $C_u = \cup\{B \in \pi' : B \cap C \neq \emptyset\}$ . A straightforward argument shows that  $C_l \subseteq C \subseteq C_u$ , and that  $\mu(C_u \setminus C_l) = \mu(\partial(C : \pi')) = \mu(\partial(C : \pi)) < \epsilon$ . It follows that  $\Theta = \{[C_l, C_u] : C \in \mathcal{C}\}$  is a collection of  $\epsilon$ -brackets covering  $\mathcal{C}$ . The cardinality of  $\Theta$  is at most  $2^{2|\pi|}$ .

## 2.2 Uniform Laws of Large Numbers

Let  $X_1, X_2, \dots$  be a stationary ergodic process taking values in  $(\mathcal{X}, \mathcal{S})$  with  $X_i \sim \mu$ . The ergodic theorem ensures that, for every  $C \in \mathcal{S}$ , the sample averages  $n^{-1} \sum_{i=1}^n I_C(X_i)$  converge with probability one to  $\mu(C)$ . For VC classes and i.i.d. sequences  $\{X_i\}$  this convergence is known to be uniform over  $\mathcal{C}$  [10]. Using Corollary 1 it is easy to show that this uniform convergence extends to ergodic processes as well.

**Theorem 2.** *If  $\mathcal{C}$  is a countable VC-class of sets and  $X_1, X_2, \dots \in \mathcal{X}$  is a stationary ergodic*

process with  $X_i \sim \mu$ , then

$$\sup_{C \in \mathcal{C}} \left| \frac{1}{n} \sum_{i=1}^n I_C(X_i) - \mu(C) \right| \rightarrow 0$$

with probability one as  $n$  tends to infinity.

**Proof:** This follows easily from Corollary 1 and the Blum DeHardt law of large numbers (c.f. [9]), which establishes that families with finite bracketing numbers have the Glivenko Cantelli property.

The uniform strong law in Theorem 2 was established in [1] using arguments similar to those for Theorem 1. Analogous uniform strong laws for VC major and VC graph classes are given in [1], while [2] contains uniform strong laws for classes of functions having finite gap (fat shattering) dimension. See these papers for a discussion of earlier and related work.

### 2.3 Uniform Mixing Conditions in Ergodic Theory

Let  $T$  be an ergodic  $\mu$ -measure preserving transformation of  $(\mathcal{X}, \mathcal{S})$ .  $T$  is said to be strongly mixing if for each pair  $A, B$  of measurable sets,  $\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$ . Theorem 1 can be applied to show that strong mixing occurs uniformly over a countable VC class.

**Proposition 1.** *If  $\mathcal{C} \subseteq \mathcal{S}$  is a countable VC-class of measurable sets, and  $T$  is a strongly mixing transformation, then*

$$\lim_{n \rightarrow \infty} \sup_{A, B \in \mathcal{C}} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| = 0.$$

**Proof:** Given  $\epsilon > 0$ , let  $\pi$  be a finite partition such that  $\sup_{C \in \mathcal{C}} \mu(\partial(C : \pi)) < \epsilon$ . Choose a natural number  $N$  such that for  $n \geq N$  and each pair  $D_1, D_2 \in \pi$ ,

$$|\mu(D_1 \cap T^{-n}D_2) - \mu(D_1)\mu(D_2)| < \epsilon \mu(D_1)\mu(D_2).$$

For every measurable set  $A$  let  $\overline{A} = \cup\{D \in \pi : \mu(D \cap A) > 0\}$  and  $\underline{A} = \cup\{D \in \pi : D \subset A\}$  be, respectively, upper and lower approximations of  $A$  derived from the cells of  $\pi$ . Note that if  $A, B$  are measurable sets satisfying  $\overline{A} = \underline{A}$  and  $\overline{B} = \underline{B}$ , then

$$\begin{aligned} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| &= \left| \sum_{D \subseteq \overline{A}} \sum_{D' \subseteq \overline{B}} \mu(D \cap T^{-n}D') - \sum_{D \subseteq \overline{A}} \sum_{D' \subseteq \overline{B}} \mu(D)\mu(D') \right| \\ &\leq \sum_{D \subseteq \overline{A}} \sum_{D' \subseteq \overline{B}} |\mu(D \cap T^{-n}D') - \mu(D)\mu(D')| \\ &< \sum_{D \subseteq \overline{A}} \sum_{D' \subseteq \overline{B}} \epsilon \mu(D)\mu(D') \leq \epsilon \mu(A)\mu(B) \leq \epsilon. \end{aligned}$$

Suppose now that  $A, B$  are sets in  $\mathcal{C}$ . Then for  $n \geq N$ ,

$$\begin{aligned}
& |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \\
&= |\mu(A \cap T^{-n}B) \pm \mu(A \cap T^{-n}\bar{B}) \pm \mu(\bar{A} \cap T^{-n}\bar{B}) \pm \mu(\bar{A})\mu(\bar{B}) \pm \mu(\bar{A})\mu(B) - \mu(A)\mu(B)| \\
&\leq 2\mu(\bar{B} \setminus B) + 2\mu(\bar{A} \setminus A) + |\mu(\bar{A} \cap T^{-n}\bar{B}) - \mu(\bar{A})\mu(\bar{B})| \\
&< 5\epsilon,
\end{aligned}$$

where the first inequality follows from the triangle inequality, and the second follows from the previous two displays. As  $A, B \in \mathcal{C}$  and  $\epsilon > 0$  were arbitrary, Theorem 1 follows.

A similar argument can be used to show that any weak mixing transformation satisfies uniform convergence over countable VC classes. A measure preserving transformation  $T$  is weak mixing if given measurable sets  $A$  and  $B$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(A \cap T^{-i}B) - \mu(A)\mu(B)| = 0.$$

**Proposition 2.** *If  $\mathcal{C}$  is a countable VC-class of measurable sets and  $T$  is a weakly mixing transformation, then*

$$\lim_{n \rightarrow \infty} \sup_{A, B \in \mathcal{C}} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(A \cap T^{-i}B) - \mu(A)\mu(B)| = 0.$$

### 3 Proof of Theorem 1

The proof of Theorem 1 follows arguments used in [1] to establish uniform laws of large numbers for VC classes under ergodic sampling, and we make use of several auxiliary results from that paper in what follows.

#### 3.1 Joins and the VC dimension

**Definition:** The join of  $k$  sets  $A_1, \dots, A_k \subseteq [0, 1]$ , denoted  $J = \bigvee_{i=1}^k A_i$ , is the partition consisting of all *non-empty* intersections  $\tilde{A}_1 \cap \dots \cap \tilde{A}_k$  where  $\tilde{A}_i \in \{A_i, A_i^c\}$  for  $i = 1, \dots, k$ .

Note that  $J$  is a finite partition of  $[0, 1]$ . The join of  $A_1, \dots, A_k$  is said to be full if it has (maximal) cardinality  $2^k$ . The next Lemma (see [6, 1]) makes an elementary connection between full joins and the VC dimension.

**Lemma 1.** *Let  $\mathcal{C}$  be any collection of subsets of  $\mathcal{X}$ . If for some  $k \geq 1$  there exists a collection  $\mathcal{C}_0 \subseteq \mathcal{C}$  of  $2^k$  sets having a full join, then  $\text{VC-dim}(\mathcal{C}) \geq k$ .*

The proof given here establishes that the approximating partition  $\pi$  is measurable  $\sigma(\mathcal{C})$ . A simple counterexample shows that it is not sufficient for the elements of  $\pi$  to belong to  $\bigcup_{n=1}^{\infty} \sigma(C_1, C_2, \dots, C_n)$ . To see this, let  $\mathcal{X} = [0, 1]$  and let  $\lambda$  be Lebesgue measure. Let  $a_1, a_2, \dots > 0$  be a sequence of numbers such that  $s = \sum_{n=1}^{\infty} a_n < 1$ . Let  $s_n = \sum_{i=1}^n a_i$  for  $n \geq 1$  and let  $s_0 = 0$ . Define  $C_n = [s_{n-1}, s_n)$  for  $n \geq 1$ . Clearly, the VC-dimension of the class  $\{C_1, C_2, \dots\}$  equals 1, since its constituent sets are disjoint. Define  $J_n = C_1 \vee C_2 \vee \dots \vee C_n$ . Then  $A_n = [s_n, 1]$  is a single element in  $J_n$  with measure  $1 - s_n > 1 - s > 0$ . Moreover, both  $A_n \cap C_{n+1}$  and  $A_n \cap C'_{n+1}$  have positive measure, so that  $\mu(\partial(C_{n+1} : J_n)) > 1 - s$  for  $n \geq 1$ .

### 3.2 Reduction to the Unit Interval

Fix a probability measure  $\mu$  on  $(\mathcal{X}, \mathcal{S})$  and let  $\mathcal{C} \subseteq \mathcal{S}$  have finite VC dimension. It follows from standard results on the  $L_p$ -covering numbers of VC classes (*c.f.* Theorem 2.6.4 of [9]) that there exists a countable sub-family  $\mathcal{C}_0$  of  $\mathcal{C}$  such that

$$\inf_{C' \in \mathcal{C}_0} \mu(C' \Delta C) = 0$$

for each  $C \in \mathcal{C}$ . An elementary argument then shows that, for every finite partition  $\pi$ ,

$$\sup_{C \in \mathcal{C}} \mu(\partial(C : \pi)) = \sup_{C \in \mathcal{C}_0} \mu(\partial(C : \pi)),$$

and we may therefore assume that  $\mathcal{C}$  is countable. Let  $\mathcal{X}_0 = \{x : \mu(\{x\}) > 0\}$  be the set of atoms of  $\mu$  and let  $\mu_0(A) = \mu(A \cap \mathcal{X}_0)$  be the atomic component of  $\mu$ . As  $\mathcal{X}_0$  is countable, it is easy to see that

$$\inf_{\pi \in \Pi} \sup_{C \in \mathcal{C}} \mu_0(\partial(C : \pi)) = 0,$$

and we may therefore assume that  $\mu$  is non-atomic.

Following the proof in [1], we make two further reductions. Let  $\lambda(\cdot)$  be Lebesgue measure on the unit interval  $[0, 1]$  equipped with its Borel subsets  $\mathcal{B}$ . Using the existence of a measure-preserving isomorphism between  $(\mathcal{X}, \mathcal{S}, \mu)$  and  $([0, 1], \mathcal{B}, \lambda)$  (*c.f.* [8]) a straightforward argument ensures that we lose no generality in assuming that  $\mathcal{X} = [0, 1]$ ,  $\mu = \lambda$ , and that  $\mathcal{C} \subseteq \mathcal{B}$  is a countable family with finite VC dimension. Using an additional isomorphism described in Lemma 6 of [1] we may further assume that each element of  $\mathcal{C}$  is a finite union of intervals.

Based on the reductions above, Theorem 1 is a corollary of the following result.

**Theorem 3.** *Let  $\mathcal{C} \subseteq \mathcal{B}$  be a countable VC class, each of whose elements is a finite union of intervals. For every  $\epsilon > 0$  there exists a finite partition of  $[0, 1]$  such that*

$$\sup_{C \in \mathcal{C}} \lambda(\partial(C : \pi)) < \epsilon.$$

**Remark:** The proof of Theorem 3 follows the proof of Proposition 3 from [1]. Beginning with the assumption that the conclusion of the theorem is false, we construct, in a step-wise fashion, a sequence of “splitting sets”  $R_1, R_2, \dots \subseteq [0, 1]$  from the sets in  $\mathcal{C}$ . At the  $k$ th stage the splitting set  $R_k$  is obtained from a sequential procedure that makes use of the splitting sets  $R_1, \dots, R_{k-1}$  produced at previous stages. The splitting sets are then used to identify finite, but arbitrarily large, collections of sets in  $\mathcal{C}$  having full join. The existence of these collections implies that  $\mathcal{C}$  has infinite VC dimension by Lemma 1.

**Proof of Theorem 3:** Suppose to the contrary that there exists an  $\eta > 0$  such that

$$\sup_{C \in \mathcal{C}} \lambda(\partial(C : \pi)) > \eta \text{ for every } \pi \in \Pi. \quad (3)$$

For  $n \geq 1$  let  $\mathcal{D}_n = \{[k2^{-n}, (k+1)2^{-n}] : 0 \leq k \leq 2^n - 1\}$  be the set of closed dyadic intervals of order  $n$ .

**Stage 1.** Let  $C_1(1)$  be any set in  $\mathcal{C}$ . Suppose that sets  $C_1(1), \dots, C_1(n) \in \mathcal{C}$  have already been selected, and let  $J_1(n) = \mathcal{D}_n \vee C_1(1) \vee \dots \vee C_1(n)$ . It follows from (3) that there is a set  $C_1(n+1) \in \mathcal{C}$  such that  $G_1(n) = \partial(C_1(n+1) : J_1(n))$  has measure greater than  $\eta$ . Let  $J_1(n+1) = \mathcal{D}_{n+1} \vee C_1 \vee \dots \vee C_{n+1}$  and continue in the same fashion. The sets  $\{G_1(n)\}$  are naturally associated with a tight family of sub-probability measures  $\{\lambda_n(\cdot) = \lambda(\cdot \cap G_1(n))\}$ . There is therefore a subsequence  $\{\lambda_{n_1(r)}\}$  that converges weakly to a sub-probability  $\nu_1$  on  $([0, 1], \mathcal{B})$ . It is easy to see that  $\nu_1$  is absolutely continuous with respect to  $\lambda$  and that

$$\nu_1([0, 1]) \geq \limsup_{r \rightarrow \infty} \lambda_{n_r}([0, 1]) \geq \eta.$$

The Radon-Nikodym derivative  $d\nu_1/d\lambda$  is well defined, and is bounded above by 1. Define the splitting set  $R_1 = \{x : (d\nu_1/d\lambda)(x) > \eta/2\}$ . From the previous remarks it follows that

$$\eta \leq \nu_1([0, 1]) = \int_0^1 \frac{d\nu_1}{d\lambda} d\lambda \leq \int_{R_1} 1 d\lambda + \int_{R_1^c} \eta/2 d\lambda \leq \lambda(R_1) + \eta/2, \quad (4)$$

and therefore  $\lambda(R_1) \geq \eta/2$ .

**Subsequent stages.** In order to construct the splitting set  $R_k$  at stage  $k$ , let  $C_k(1)$  be any element of  $\mathcal{C}$ , and suppose that  $C_k(2), \dots, C_k(n)$  have already been selected. Define the

join

$$J_k(n) = \mathcal{D}_n \vee \bigvee_{j=1}^{k-1} R_j \vee \bigvee_{i=1}^n C_k(i). \quad (5)$$

By (3) there exists a set  $C_k(n+1) \in \mathcal{C}$  such that  $G_k(n) = \partial(C_k(n+1 : J_k(n)))$  has measure greater than  $\eta$ . This process continues as in stage 1. As before, there is a sequence of integers  $n_k(1) < n_k(2) < \dots$  such that the measures  $\lambda(B \cap G_k(n_k(r)))$  converge weakly to a sub-probability measure  $\nu_k$  on  $([0, 1], \mathcal{B})$  that is absolutely continuous with respect to  $\lambda(\cdot)$ . Define  $R_k = \{x : (d\nu_k/d\lambda)(x) > \delta\}$ .

**Construction of Full Joins.** Fix an integer  $L \geq 2$ . As the measures of the sets  $R_k$  are bounded away from zero, there exist positive integers  $k_1 < k_2 < \dots < k_L$  such that  $\lambda(\bigcap_{j=1}^L R_{k_j}) > 0$ . Suppose without loss of generality that  $k_j = j$ , and define the intersections

$$Q_r = \bigcap_{j=1}^{L-r} R_j$$

for  $r = 0, 1, \dots, L-1$ . Note that  $Q_0 \subseteq Q_1 \subseteq \dots \subseteq Q_{L-1}$ . We show that there exist sets  $D_1, D_2, \dots, D_{L-1} \in \mathcal{C}$  such that, for  $l = 1, \dots, L-1$ ,

- (i) the join  $K_l = D_1 \vee D_2 \vee \dots \vee D_l$  has cardinality  $|K_l| = 2^l$ , and
- (ii)  $B^o \cap Q_l$  is non-empty for each  $B \in K_l$ , where  $B^o$  denotes the interior of  $B$ .

We proceed by induction, beginning with the case  $l = 1$ . Let  $x_1$  be a Lebesgue point of  $Q_0$ , and let  $\epsilon = \eta/2(\eta+2)$ . Then there exists  $\alpha_1 > 0$  such that the interval  $I_1 \triangleq (x_1 - \alpha_1, x_1 + \alpha_1)$  satisfies

$$\lambda(I_1 \cap Q_0) \geq (1 - \epsilon)\lambda(I_1) = 2\alpha_1(1 - \epsilon). \quad (6)$$

It follows from the last display and the definition of  $R_L \supseteq Q_0$  that

$$\nu_L(I_1 \cap R_L) = \int_{I_1 \cap R_L} \frac{d\nu_L}{d\lambda} d\lambda > \frac{\epsilon\eta}{2} \lambda(I_1 \cap R_L) \geq \alpha_1(1 - \epsilon)\eta. \quad (7)$$

Let  $\{n_L(r) : r \geq 1\}$  be the subsequence used to define the sub-probability  $\nu_L$ . As  $I_1$  is an open set, the portmanteau theorem and (7) imply that

$$\liminf_{r \rightarrow \infty} \lambda(I_1 \cap G_L(n_L(r))) \geq \nu_L(I_1) \geq \nu_L(I_1 \cap R_L) > \alpha_1(1 - \epsilon)\eta.$$

Choose  $r$  sufficiently large so that  $\lambda(I_1 \cap G_L(n_L(r))) > \alpha_1(1 - \epsilon)\eta$  and  $2^{-n_L(r)} < \eta\alpha_1/8$ .

We require the following lemma from [1].



**Lemma 2.** *There exists a cell  $A$  of  $J_L(n_L(r))$  such that  $A \subseteq \partial(C_L(n_L(r) + 1) : J_L(n_L(r)))$ ,  $A \subseteq I_1$  and  $\lambda(A \cap Q_1) > 0$ . Moreover,  $A$  is contained in  $Q_1$ .*

Let  $D_1 = C_L(n_L(r) + 1) \in \mathcal{C}$ , and let  $A$  be the set identified in Lemma 2. By definition of the boundary,  $\lambda(A \cap D_1) > 0$  and  $\lambda(A \cap D_1^c) > 0$  and therefore  $\lambda(Q_1 \cap D_1) > 0$  and  $\lambda(Q_1 \cap D_1^c) > 0$  as well. As the Lebesgue measure of the boundary  $\overline{D_1} \setminus D_1^o$  of  $D_1$  is zero, assertion (ii) above follows.

Suppose now that we have identified sets  $D_1, \dots, D_l \in \mathcal{C}$ , with  $l \leq L - 2$ , such that (i) and (ii) hold. Let the join  $K_l = \{B_j : 1 \leq j \leq 2^l\}$ , and for each  $j$  let  $x_j \in B_j^o \cap Q_l$ . Select  $\alpha_{l+1} > 0$  such that for each  $j$  the interval  $I_j \triangleq (x_j - \alpha_{l+1}, x_j + \alpha_{l+1})$  is contained in  $B_j^o$  and satisfies

$$\lambda(I_j \cap Q_l) \geq (1 - \epsilon)\lambda(I_j) = 2\alpha_{l+1}(1 - \epsilon).$$

To simplify notation, let  $\kappa = L - l$ . Let  $\{n_\kappa(r) : r \geq 1\}$  be the subsequence used to define the sub-probability  $\nu_\kappa$ . For each interval  $I_j$ ,

$$\liminf_{r \rightarrow \infty} \lambda(I_j \cap G_\kappa(n_\kappa(r))) \geq \nu_\kappa(I_j) \geq \nu_\kappa(I_j \cap R_\kappa) > \alpha_{l+1}(1 - \epsilon)\eta,$$

where the last inequality follows from the previous display, and the fact that  $Q_l \subseteq R_\kappa$ . Choose  $r$  sufficiently large so that  $\lambda(I_j \cap G_\kappa(n_\kappa(r))) > \alpha_{l+1}(1 - \epsilon)\eta$  for each  $j$ , and  $2^{-n_\kappa(r)} < \eta \alpha_{l+1}/8$ .

By applying the Lemma 2 to each interval  $I_j$ , one may establish the existence of sets  $A_j \in \partial(C_\kappa(n_\kappa(r) + 1) : J_\kappa(n_\kappa(r)))$  such that  $A_j \subseteq I_j \subseteq B_j^o$ ,  $\lambda(A_j \cap Q_{l+1}) > 0$ , and  $A_j \subseteq Q_{l+1}$ . Let  $D_{l+1} = C_\kappa(n_\kappa(r) + 1) \in \mathcal{C}$ . Arguments like those for the case  $l = 1$  above show that for each  $j$  the intersections  $A_j \cap D_{l+1}^o$  and  $A_j \cap (D_{l+1}^c)^o$  are non-empty, and the inductive step is complete. Given any two dyadic intervals, they are disjoint, intersect at one point, or one contains the other. Therefore, among the sets  $D_1, \dots, D_{L-1}$ , at most one can be a dyadic interval; the remainder are contained in  $\mathcal{C}$ .

## Acknowledgements

The authors would like to thank Ramon van Handel for pointing out an oversight in the proof of Corollary 1. The work presented in this paper was supported in part by NSF grant DMS-0907177.

## References

- [1] ADAMS, T.M. and NOBEL, A.B. (2010) Uniform convergence of Vapnik-Chervonenkis classes under ergodic sampling. *Annals of Probability* **38:4** 1345–1367.
- [2] ADAMS, T.M. and NOBEL, A.B. (2010) The gap dimension and uniform laws of large numbers for ergodic processes. arXiv:1007.2964v1 [math.PR]. Submitted for publication.
- [3] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed., Wiley, New York. MR1324786 (95k:60001)
- [4] DEVROYE, L. and GYÖRFI, L. and LUGOSI, G. (1996) *A Probabilistic Theory of Pattern Recognition* Springer.
- [5] DUDLEY, R.M. (1999) *Uniform Central Limit Theorems* Cambridge Univ. Press, New York. MR1720712 (2000k:60050)
- [6] MATOUSEK, J. (2002) Lectures on Discrete Geometry. *Graduate Texts in Mathematics* **212** Springer, New York. MR1899299 (2003f:52011)
- [7] POLLARD, D. (1984) *Convergence of Stochastic Processes* Springer, New York. MR0762984 (86i:60074)
- [8] ROYDEN, H.L. (1988) *Real Analysis*, 3rd ed. Macmillan Publishing Company, New York. MR1013117 (90g:00004)
- [9] VAN DER VAART, A.W. and WELLNER, J.A. (1996) *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671 (97g:60035)
- [10] VAPNIK, V.N. and CHERVONENKIS, A.YA. (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280. MR0627861 (83d:60031)